DOCUMENT RESUME

ED 171 377 PS 010 553

AUTHOR	Dykes, Mary K.; And Others
TITLE	Florida Educational Research and Development Council, Inc. Assessment and Evaluation Strategies for Early Childhood Programs. Research Bulletin.
INSTITUTION	Florida Educational Research and Development Council, Gainesville.
PUB DATE	79
NOTE	70p.; For supplement to this document, see PS 010 554
AVAILABLE FROM	FERDC, 2266 2nd Street, Fort Myers, Florida 33901 (Paper, $2.00)
JOURNAL CIT	Florida Educational Research and Development Council, Inc., Research Bulletin; v12 n3 Winter 1979
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Administrative Personnel; Early Childhood Education; Educational Assessment; *Educational Testing; *Evaluation Methods; *Evaluation Needs; Preschool Children; Preschool Evaluation; Preschool Teachers; Program Development; *Program Evaluation; Relevance (Education); *Testing Problems; *Test Selection
IDENTIFIERS	*Florida

ABSTRACT

This brief monograph, intended for preschool teachers and administrators, explains various aspects of the assessment and evaluation of children and early childhood education programs. The value of evaluation procedures for program development is noted. The different forms that research designs and tests can take are explained. It is emphasized that the persons responsible for evaluation must choose tests and research designs which do measure what they are supposed to measure; examples are used to show how this can fail to happen if care is not taken. The information the examiner should have about each child prior to testing is indicated. A checklist of conditions for testing (e.g., good lighting and quiet) and federal requirements for evaluation are included. (BH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED1713777

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

FLORIDA EDUCATIONAL
RESEARCH AND DEVELOPMENT COUNCIL

RESEARCH BULLETIN

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Florida Educational Res. + Dev. Council
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

PS010553

ASSESSMENT AND EVALUATION STRATEGIES FOR EARLY CHILDHOOD PROGRAMS

Volume 12 Winter, 1979 Number 3

FLORIDA EDUCATIONAL
RESEARCH AND DEVELOPMENT COUNCIL, INC.

RESEARCH BULLETIN

ASSESSMENT AND EVALUATION STRATEGIES FOR EARLY CHILDHOOD PROGRAMS

by

Mary K. Dykes, Ph.D.
Associate Professor
University of Florida
Gainesville, Florida

and

Anne M. Strickland, M.Ed.
Program Training Specialist
Alachua County School System
Gainesville, Florida

and

Dolores D. Munyer, M.Ed.
Project Director
Alachua County School System
Gainesville, Florida

Published by
Florida Educational Research and Development Council, Inc.

Additional copies of this book may be obtained from
FERDC
2266 Second Street, Ft. Myers, Florida 33901

Individual Copies $2.00
Annual Subscription (4 issues) $8.00

10% discount on an order of 5 or more.
Order of less than $20.00 must be accompanied by check or money order.

PS 010553

Foreword

In the most recently proposed Florida list of "Elementary and Early Childhood Essential Teaching Competencies", you will find competency #52: "Utilizes a variety of formal and informal diagnostic and evaluative procedures that are developmentally appropriate to assess the needs, interests, abilities, and achievements of the child." This competency is projected as one of the essential skills needed by pre- and inservice teachers. In this Bulletin three people, Mary K. Dykes, Anne M. Strickland, and Dolores D. Munyer, who have constant contact with teachers and administrators in a variety of early childhood programs, have summarized the current thinking in evaluation and testing applicable to preschool

and early childhood programs. Their excellent, practical presentation provides teachers and administrators with the background knowledge which is needed to perform a quality evaluation. Mary K. Dykes, Anne M. Strickland, and Dolores D. Munyer are to be commended for a job well done.

W. F. Breivogel
Guest Editor

If you are interested in a listing of early childhood instruments you may want the supplement to this Bulletin (Assessment and Evaluation Instruments for Early Childhood Programs by Mary

K. Dykes, Anne M. Strickland, and Dolores D. Munyer). It may be obtained by sending a check or money order for $2.00 to:

FERDC
2266 2nd Street
Fort Myers, Florida 33901

Contents

Introduction.....1

Purpose of Monograph.....3

Federal Requirements for Evaluation.....7

Types of Evaluation Designs.....10

Testing Do's and Don'ts.....29

Test Selection Guidelines.....35

Checklist for Establishing a Testing Environment.....49

Information to Know Before Beginning to Test a Child.....57

Introduction

During the past three years a recurring need has been brought to the attention of the authors by school district administrators and teachers concerning the difficulties in assessing and evaluating preschool and other early childhood education programs. The demand for quality of information collected concerning the effectiveness of new developments in early childhood education has increased significantly. At the same time, difficulties in developing viable strategies to gather good, usable data has increased also.

Programs in which effectiveness in cost and in

learning can be demonstrated are usually the programs that local school boards, or state agencies consider for full scale or partial implementation. It is to an administrator's and staffs' advantage to investigate, examine and then develop sound assessment and evaluation strategies that are appropriate for and sensitive to the outcomes of the particular educational program. This monograph has been developed to help meet these needs and to suggest sound strategies to assist in the development of assessment and evaluation plans for various early childhood education programs.

Purpose of Monograph

This monograph represents an overview of selected educational assessment and evaluation strategies for use with very young children. The purpose of this text is to help educators plan and develop evaluation strategies for three-, four- and five-year old children in the areas of cognitive-intellectual development, language development, perceptual motor development, and social-emotional development.

In order to have meaningful operational curriculum objectives, a plan or design for the type of assessment program must first be devised. Assessment results then become a primary information

source for the derivation of curriculum objectives for the program and for each child. Short term feedback information and long range evaluation procedures designed to assess the general effectiveness of the intervention program should be closely aligned with the behavioral objectives and original needs as determined through assessment. The results from the assessment data are valuable reference points for the selection of final measurement formats. It is wise to note that the final evaluation design and methodology, in particular, presents a series of critical considerations. To a large extent, the potential success of any early learning program depends directly upon the virtues and faults of the evaluation design.

This monograph provides basic information and strategies which should be considered before attempting to implement an evaluation design in programs for very young children. School board districts, schools and agencies that are implementing "new" programs or trying to prove that "old" programs are successful must be able to provide evidence that their program is valuable and that it is of educational benefit to the children for whom it was designed. This monograph addresses the needs expressed to the authors by school district personnel regarding their difficulty in documenting the effectiveness of early childhood educational programs. The authors observed that these difficulties in documentation of early learning programs were

related primarily to the type of assessment measures being used and the type of evaluation design being selected.

Therefore, this monograph will report on pertinent information needed in developing assessment and evaluation strategies for use in early childhood education programs.

Federal Requirements for Evaluation

The staff of the Office of Civil Rights (OCR) has developed a set of criteria to be applied to all evaluations of pupils in programs for which federal funds are received. This includes all programs that receive money through the state department of education as well as those programs that are funded directly by a federal source.

The seven points in the criteria as reported by Bersoff and Ysseldyke are:

1. No placement or denial of placement can

Source: Bersoff & Ysseldyke. Paper presented at ACLD Conference, Washington, D.C., 1977.

occur without a full and individual evaluation of each child's special educational needs.

2. Any evaluation must be in the primary language of the student.
3. All evaluation devices must be recommended by their producer for the specific purpose for which the school proposes to use them and administered only by trained personnel.
4. All evaluation devices must be properly and professionally validated for the specific purpose for which the school proposes to use them.
5. Any assessment must contain more than tests designed to provide a single general intelligence quotient.

6. No one test or means of evaluation can be used as the sole criterion for placement.
7. Evaluation, other than tests, or such attributes as adaptive behavior, physical condition, and cultural background must be administered. Finally, the proposed regulations require an annual re-evaluation of each child's progress and needs.

All persons responsible for evaluation should be sure that the criteria as established by the OCR are adhered to strictly and are incorporated into all conceptual strategies of testing children and of program evaluation.

Types of Evaluation Designs

An adequate evaluation must include valid and reliable measures. Examples of measures which may be valid for use in early childhood programs are discussed in the "Test Selection" unit of this monograph. The assessment measures selected should be related directly to the goals and objectives of the program, and should have a logical, precise relationship to what is being measured. There are many types of measures that can be acceptable if the concepts of validity and reliability are understood in relation to what information is desired. For instance, there are measures that are obtained directly from the students which may include achievement scores or scores on

attitude or self-concept scales. Direct observational measures of change include measures of such things as absenteeism, indicators of parent involvement or observation of student behaviors in a classroom.

Several designs which may be used to demonstrate that a program is successful will be discussed briefly in this section. It should be noted that the ideal design is the classical experimental design, but, this type of design is extremely difficult, if not impossible, to use in a classroom setting where all possible variables can not be controlled. Most other evaluation designs derive their credibility from the extent to which they provide reasonable substitutes for the

elements of the classical experimental model.

One example of evaluation design which may be used is the nonequivalent control group. This design varies from the classical experimental design in the way the estimate is generated for scores for children in the group without-intervention. If post test scores of a randomly assigned control group are not available, one can derive estimates from the following sources as reported by Tallmadge (1977):

*a. a highly similar but nonequivalent control group (the children don't have the same

*Adapted From: Tallmadge, G. Kasten. Ideabook, Washington, D.C., National Institute of Education, October 1977.

teacher, or attend school in the same building, but are alike in basic characteristics).

b. historical data (how a similar group of children performed last year).

c. a comparison group formed by splitting a group of children into a treatment and a comparison group based on a cutoff score on a pre test (each group would contain the same number of children who scored at a specific level on the test).

d. national, regional or local normative data (comparison to a group of children of the same age and general characteristics in a similar program located somewhere else).

For the purposes of this monograph no attempt will be made to describe sound evaluation designs that can be implemented in each of the above conditions. Instead designs will be described that are more commonly used, especially in norm referenced evaluation.

Designs that are frequently used to study the performance of groups of children include:

1. Nonrandom selection of critical and experimental groups - In this design children are assigned for some justifiable reason to a group. At times children may be grouped for study based on variables such as number of years in the program, socio-economic level, chronological age, test scores, etc.

When using this design, evaluators must be careful to be sure that assignment to a group is based on a justifiable reason in terms of the type of data to be gathered and the basic question for which information is sought. A major problem with this design is that the control and experimental groups may not be comparable.

2. Nonrandom selection with correction by analysis of covariance. It is possible for a design specialist to assist the program staff in using this method for data collection. It has numerous advantages since the staff can retrieve data using different variables as the covariant or constant. It is difficult to

know the magnitude of bias introduced into the data when using this method. It is suggested that the school board research specialist or a professor of educational design and evaluation (in the educational psychology or foundations department) from a nearby college of education should be requested to help develop this design if it is selected for use.

3. Matched sample design with control group drawn from another population. The program staff may select this design if it can be justified as to why performance of children in this setting and program should be compared with the performance of children in the selected group. The control

group may be a national, state, or local population with whom it would logically be expected that the children in the current program would be alike or different.

Using this method the performance of children in the pre-school program can be compared with that of children in the district or a similar district who did not participate in the intervention program. Commonly used matching variables in this design include: race, sex, age, and socio-economic status.

4. Design using pretest scores as age norms. This procedure may be appropriate when the group under study contains children in a

relatively wide age range (e.g. 2-5 years). Children are divided into groups based on an age range interval (e.g. 6 months). A mean chronological age and a mean pretest score are calculated for each group. Scores of children in the control groups are then compared to the group being studied using the same type data. In this design the pretest scores of children who start the program but drop out for some reason or who join the program after the pretest can still be included in the data. Comparisons are not made of scores of an individual child's pre and post intervention performance.

Problems Introduced in Evaluation Designs

In almost every case it can be argued that the design selected has problems - that it is biased, non-sensitive, inappropriate, etc. If the program staff are aware of the most frequent ways that errors are committed in the selection of a design, then perhaps they can use these pitfalls as criteria to evaluate the design they have selected. Frequently reported errors in the design of a project evaluation model include:

1. Using too simple a design such as the sole use of pre-post test data. Indicating that a child, after a year of school, has gained score points proves nothing about the program.

the child was in. As a function of maturation and exposure to the world around him/her a child would be expected to obtain a higher score on a measure after nine months.

2. Stating that when normative data is gathered on disadvantaged children, that, just because they show greater skill gain than the number of months they have been in the program, they have moved closer to the national norm. This is not necessarily true. The difference is created based on the way grade level or grade equivalent scores are constructed by test publishers.

3. Making exaggerated claims that cannot be supported by the data - e.g. that the program

made a difference in the skill levels of participants when the only data collected concerned parents perceptions of their child's change. The claim cannot be supported by the data. It may or may not be true.

4. Using only one score to select children for program and for pre-intervention data. Students should be tested on at least two measures or retested on a single measure to establish baseline data. This is important since the measurement of error for a given test is frequently within ± 10 points of the obtained score. Especially when a child scores very low or very high it is important to have data from several measures which all

support the trend that the child is indeed far below or above the performance expected for a child of his age, grade, etc. When re-tested most children tend to score closest to the mean score than they did previously, especially if their first score was far from the mean. This is the construct of regression.

5. Using different dates for testing and then comparing performance. If one group is tested in September or within the first month of school then the comparison group should be tested during the same period.

6. Using inappropriate levels of tests. As a general rule, children should be tested using

the test form for which their achievement level is near the mean. For example, the chance of obtaining accurate data for a normal five year old is greater on a test that is designed to measure nursery through first grade level skills than on one designed to measure kindergarten to second grade level skills. It is most important that tests be selected based on achievement levels of the child, not chronological age and that the achievement level of the child be in the middle (or near the mean) of the skills tested on the test level selected.

7. Using data that has a bias. For most designs it is important that data for pre and post

assessment be present for each child. It is possible that children for whom there is missing data have characteristics in common that could introduce bias into the evaluation. For instance, if children with lower scores were dropped from the program before post test data were collected or if higher scoring children were graduated before post test data were collected, it is easy to see that the post test scores would be greatly changed.

8. Using non-comparable groups. Bias can easily be introduced into the data if the groups being studied are not alike on important characteristics. Selection of groups based on a random process is the best way to control

for bias.

9. Using inappropriate scores for comparison. Raw scores should not be compared. The school board research specialist or a specialist in research design should be asked to help select the type of scores which will be most appropriate for making the desired comparisons.

10. Matching a control group to a selected group. It is almost impossible to match two groups. Designs should be avoided that require matching. Random assignment or other statistical assignment procedures are greatly desired over matching.

11. Collecting data carelessly. Several problems

are related to this issue -

a. not recording data appropriately while testing.

b. use of an examiner known to the children who may influence the scores recorded for children based on factors other than strict interpretation of responses. When possible, one person should do all testing and should not be involved with the children for anything other than testing.

c. use of several examiners for whom inter rater reliability has not been established.

If more than one tester is required

then they should be trained by the same individual and should practice giving the test until they are consistently within 95% of each other in establishing scores for the same performance.

12. Using different instruments for pre/post testing. Scores from one instrument should not be compared with those from another for pre/post measures. There is no way to know if exactly the same skills and levels of skills were tested even if both tests are measures of language for 5 year olds, etc.

13. Implying causes from data. When test scores are indicative of low or high performance it

is almost second nature to speculate as to the cause - e.g. "Sure he's low - he never gets to go anywhere or he has no books at home, etc." For most measures there is no way to know specifically why a child scores as he does. All that can be said is that at this time, this is his score.

With all these pitfalls or cautions, it seems almost impossible to design any acceptable evaluation model, however this is not true. If an evaluation model is selected which is comprehensive and which minimizes the above pitfalls to the greatest degree possible, then it is probably as good as is possible for the conditions under which the program must operate.

Testing Do's and Don'ts

Almost all measures help teachers to gain information about some child at some time and under some conditions; however, there is no one single measure or process of evaluation which is right for all children all the time. If educators are knowledgeable about testing, they will insist that a great deal of thought by all concerned professionals and parents be given to developing the optimal evaluation battery possible in order to gain specific information regarding the child's skills, interests and/or perceptions. The following points are intended as guiding principles in selecting and utilizing tests:

1. DO decide exactly what it is you desire to learn about the child, e.g. Does the child know the letters of the alphabet? Not why doesn't Johnny read? Try to isolate academic and behavior or performance questions for which data is sought, e.g. Does Sam know the 6 basic colors? What behaviors interfere with Sam's learning when he is supposed to be working independently?
2. DO try to establish specifically what the child can do in each skill area in question so that programming can be initiated at the exact level of a child's ability in order to help the child learn the next higher level of the skill.

3. DO believe in and seek out measures, interviews, and observations obtained by other professionals, and by family members concerning the child in order to gather comprehensive information about the child's performance, unique abilities, and preferences before beginning to test or to select testing instruments.
4. DO plan ahead. Know what materials, space, and time will be needed to gather desired information from the child before beginning to test.
5. DO practice giving all tests before beginning with the child. Even if you've given the test before plan a "test-run" before administering the test to a child for diagnostic purposes.

6. DO NOT assume that all tests measure what they say they do just because it states in the manual that it is a test of intelligence, etc. Realize that few persons agree in "what is intelligence" and that there are many, many types of measures of intelligence, achievement, aptitude, etc. The content format and level of information required of the child vary widely even in tests that are supposed to measure the same skill or ability.
7. DO NOT believe that a test measures motivation or that test scores are absolute. "Overachievers" don't exist unless we assume an absoluteness about a test score. No one can perform better than he/she can perform!

8. DO NOT believe that a child's performance on a test today will be the same as tomorrow. For many measures the true score of a child will be somewhere within ± 20 points of the score obtained on this testing.
9. DO NOT assume that just because members of a minority or sex group consistently score lower on a measure that the measure is biased.
10. DO NOT assume that an achievement test tests all a child knows about an area. It depends on what was asked and how it was asked. Look for tests that test fact and knowledge and reasoning ability in an area.
11. DO NOT assume that scores in a test battery

or on subscales of a test can be compared. Often points are tallied as different types of scores that can't be compared, e.g., intelligence quotient, social quotient, grade placement, etc. A subscale may contain many more items and be a much more sensitive measure than the next subscale that has fewer items.

12. DO NOT assume that projective tests or tests of personality provide answers to why a child doesn't read, have friends, etc., or that they reveal some deep seated unchanging trait about a child. As a general rule these tests do not help teachers develop education programs for children.

Test Selection Guidelines

The selection of the most useful test or test battery is dependent upon what information is desired for the group or individual child. If the goal is to gain general data or to show change in the learning skills of children over time, then there are several guidelines to follow in selecting the measures to be used.

I. One of the first decision to be made is whether to collect screening level or comprehensive level data. Screening measures are designed to give an over-all or "ball-park" picture of the general level and abilities of the child. Usually, screening measures take

a shorter amount of time to administer than the more comprehensive or in-depth measures. Screening measures are appropriate to use 1) to identify children who may need further or more in-depth testing, 2) to gather data on general change in performance occurring within the group, 3) to estimate if a certain type of program is making a difference in childrens' performance or in a specific child's gain in skills.

It is difficult to program from a screening level instrument. If an individual plan or remedial program is being established for a child, additional testing beyond the screening level is needed. Some of the more frequently

used screening level measures for early childhood and pre-school programs include: Alpern-Ball Developmental Profile, Denver Developmental Screening Test, Developmental Activities Screening Inventory (DASI), and the Pre-School Attainment Record (PAR).

Measures that are more comprehensive than the screening scales provide more detailed information about a child's abilities - the skills that are mastered, are emerging, or are not present. These measures are designed to serve as the basis for program development for an individual. They take longer to administer than screening measures and must usually be given to one child at a time. More comprehen-

sive measures frequently used in pre-school and early childhood education programs include: Basic School Skills Inventory (BSSI), Kraner Pre School Math, Portage, Learning Accomplishment Profile (LAP), and the Wabash Guide.

II. Another guideline to remember when selecting an appropriate measure is that the test design, the items selected, and the presentation of the test should not discriminate against the child. For example, if an items asks "what is this?" for an illustration of a woman's handbag and the manual states the answer is "pocketbook" but does not allow the answers "wallet" or "purse" then children who use the

latter two words will miss the item but are correct in their response. Another example is the test that asks for a response such as "what is the opposite of ____?" If the child has not been taught the concept of "opposite" he fails the test. Had the item been presented in a different way he might have answered correctly.

Tests should be selected that are a direct measure of what information is desired. For instance, if we want to know if a child can copy his first name, we should ask the child to do exactly that - not to copy abstract forms, pictures, etc. If we want to know if a child can identify letter shapes, we should

ask him to match or point to letters, not stars, swiggles or three dimensional shapes.

III. In order for scores to be as accurate as possible, select a measure for which the child's chronological age is in the middle of the age group for whom the measure is intended. For instance, when testing a four year old the best measure would have items for two year olds and six or seven year olds on it. If tests include too wide an age range they usually become generalized screening tools offering few evaluation items at each chronological age level.

IV. If the examiner desires general data for child and or program evaluation, a comprehensive battery of tests should be administered. For pre school age children such a battery should include a measure for each major area of development - physical, motor, intellectual, language, social/emotional, and activities of daily living. For young school age children such a battery should include measures of language, academic, aptitude and social/emotional skills.

Comprehensive batteries may be at the screening or pinpoint levels of testing. Preschool level screening batteries such as the Denver or the Alpern-Ball are available as are the

more pinpoint level batteries of the Portage, Wabash, and LAP. For children with kindergarten and primary level skills, academic measures in reading, math, writing and spelling should be included. Depending on the child's level of functioning, the Basic School Skills Inventory, Kraner PreSchool Math, or Basic Concepts Inventory may be appropriate for inclusion in a battery.

The one sub scale that has been identified in the research literature as most closely related to general intellectual functioning is that of language. Staff members must understand that a measure of language such as the Peabody Picture Vocabulary Test (PPVT) may

provide an insight into the child's receptive language skill but does NOT measure all the skills included in an intelligence measure. Staff members should be careful not to discriminate against children by converting a language score to an intelligence quotient just because there is frequently a strong correlation between language and intelligence scores or because the possibility of such a conversion is discussed in the manual.

V. Depending on the information desired about the child, select the appropriate norm referenced, domain referenced, or teacher made instrument. Each type of instrument is helpful in learning about a child's abilities, but

each has a different use and value as well as set of limitations.

Norm-referenced tests are those frequently referred to as standardized. They are ones which may be used to compare the child's performance to those of a large population of children of the same age. If a child's scores are low on such a measure it may mean many different things such as the child had not been taught or exposed to the material, the child did not understand the directions or was distracted during the testing, etc. A low score on a standardized test, especially one administered to a large group of children at one time, may or may not be indicative of

learning problems. Some of the more frequently used standardized tests are the Bayley Scales, the Weschler Intelligence Scale for Children - Revised, the Stanford-Binet, the Metropolitan test of readiness, and the Iowa, Stanford, or California achievement or mental maturity tests.

Domain referenced or criterion referenced tests are those based on a hierarchy of skill development. Although hierarchies have usually been standardized and norm referenced, the domain referenced test is used in a different manner than tests in the cateory above. Most often these tests are used not for placement or for required scores but to learn of

the highest level of skill attained by the child. They also are used to determine change which has occurred in the child's learning since he was last observed. Domain referenced measures may have the appearance of a checklist of objectives or may appear in the same format as tests described in category one above. Examples of domain or criterion referenced tests are the developmental measures such as the Gesell, Portage; and the Oregon diagnostic tests, as well as the Prescriptive Reading or Math Inventory.

Teacher made measures are ideal for evaluating if the child has understood what has been presented or if the child can perform a specific

skill. Pitfalls to be avoided in teacher made tests are: 1) teaching at one level of skill such as the comprehension level but testing at a higher level such as the evaluation level (e.g. this is a cat - If you were walking home and saw this animal in the road would you be afraid?) 2) using a format or set of directions that is unfamiliar to the child, and 3) requiring the child to copy from the board or follow verbal directions when he doesn't have the skill.

In general, all types of tests can be helpful in gaining information concerning a child's performance and/or a program's merit. Staff members must be careful in selecting a

specific type of test or in developing a test battery to make sure that it meets their needs better than any other type of measure available and that all the data collected are as useful as possible in meeting the specific objectives for testing.

Checklist for Establishing a Testing Environment

There are two primary types of tests for which educators must structure the best possible testing environment - group tests and individual tests. For group tests the responsible party should check to see that:

A. The area selected is the least distracting one possible - free from kitchen, playground, and street noises.

B. The area is well lighted. Few large areas (such as cafeterias) have enough direct light where the children are working. Additional lighting may have to be obtained.

C. The children are comfortable physically -

NO child should have to sit through a testing session without his/her 1) feet resting comfortably on the floor, 2) arms supported at a comfortable height, and 3) back being supported.

D. The materials for testing are plentiful and that extra supplies are available - that materials are readily available to the examiner but are not distracting to the child.

E. The children have enough adult supervision to provide needed help with following directions and with behavior management - No one adult should try to discipline the whole group by yelling, moving around, whistling, etc.

For individual tests there are two types of environments which need to be well planned for testing.

For observation -

A. The child should have the normal freedom of territory, activity, and behavior so that the desired behavior or skill may be observed.

B. The observer should be able to see the child at all times but not be obvious, conspicuous or disruptive while observing.

C. The observer should have the use of a desk or table for papers.

For individual testing -

A. The furniture should fit the child - that is,

his feet should be comfortably on the floor with his back well supported. Extended arms should rest comfortably on the table. All legs of the chair should be the same length to prevent rocking. An additional table should be near the examiner, out of reach of the child, for placement of materials.

B. The area should be well lighted. Make sure that shadows are not on the materials or on the examiner's face.

C. The area should be free of distractors - interruptions, noise, toys, distracting displays, etc.

D. The child should face a blank wall or a non-stimulating area. Be careful that the child is not facing a window, mirror, light, or

glass partition, etc.

E. The child and the examiner should have time together before beginning testing to get to know one another, to establish rapport, and for the examiner to gain helpful information re: interests, concerns, etc.

Establishing an appropriate area for testing is critical if the data collected are to be of value in planning for the child. The tasks of organizing the area, of establishing rapport with the child, and of gaining information from parents and professionals before testing are as important in understanding the child as is any testing data.

In addition to the physical environment and

establishing rapport, other points to consider before testing are:

1. Be sure the child is feeling well and rested.
2. Tell the child how he is to respond (to point to a picture, draw a picture of ____, pick up ____) to what you say or ask. Repeat directions whenever necessary.
3. Offer praise for cooperation, for trying, not for correct responses (don't say, "that's correct", "right"; instead say, "I like the way you're thinking before you answer," or "you're really trying hard" or "you're doing a good job of looking at all the pictures before you answer", "I'm really enjoying being with you. You're doing such a nice job",

etc.)

4. Proceed at the child's pace. Some children answer quickly - do not ask them to "try again" if incorrect unless the test manual states you should. Some children are shy or simply respond more slowly than others. Encourage responding by saying "Do you want to try"? or "Do you understand what you should do"?. Never say "hurry up, we have a lot to do". Children from an Asian or Spanish culture may have been taught at home not to answer an adult quickly. If no response is made at all, move on to the next question. It's helpful to reflect on you own test-taking background to appreciate how the child may feel.

5. If a child is just not responding at all, he is not ready for formal testing in this environment with this examiner and any score will be invalid. Try retesting at a later date after more rapport has been established. Perhaps the examiner will need to start with a lower level test or lower level items.
6. Start with items that the examiner is sure the child will know. If the child becomes frustrated the examiner may want to drop back and present some easier items to that the child will feel like he/she can succeed.

Information to Know
Before Beginning to Test a Child

A well thought out plan or procedure is most helpful in gaining the most information on the child in the shortest amount of time. In addition to observing the child in the home and/or classroom area before beginning formal testing the following data should be gathered.

1. Specific questions and needs of the staff relevant to data to be collected.
2. Parental opinion and observations concerning the child's performance.
3. Professional opinions available from others who are providing any related services to

the child and/or family.

Once this data is gathered then the examiner should determine specifics about the child's performance that will help the examiner to establish rapport and to proceed as quickly as possible. Points which the examiner should consider include:

1. Does the child have any known or suspected hearing or visual problems? Has the child ever been examined by a hearing and/or visual professional? Are reports available concerning these exams? If not, this data must be gathered before testing.
2. How does the child communicate? Does he have

articulation or language patterns that make it difficult to understand him? Does he have an unusually long "thought time" before he answers? Does he speak a different language at home than he does in school? Will the child's abilities probably be demonstrated better if he is tested in a language other than English?

3. Does the child have any type of physical disabilities? Does he take medications? Does he have certain periods when he is unusually tired or restless? Does the child become distracted easily?
4. What does the child like to do, to see, to eat, etc. that could be used to help motivate

the child if interest wanes during testing?

5. How long can I expect the child to maintain interest in the testing material?
6. Are there types of tests that the child should not be expected to perform due to lack of skill, frustration or responses required?

A MAJOR GUIDING PRINCIPLE MUST BE - IF I COMPLETE THE TESTING AS PLANNED AND INCLUDING DATA I ALREADY HAVE, WHEN I FINISH, WILL I HAVE THE SPECIFIC INFORMATION NECESSARY TO ANSWER THE ACADEMIC AND BEHAVIOR QUESTIONS I HAVE CONCERNING THE CHILD'S PERFORMANCE? WILL I BE ABLE TO DISCUSS THE CHILD'S ABILITIES AND SKILL LEVELS WHEN I FINISH? OR WILL I ONLY KNOW WHAT THE CHILD CANNOT DO?

Bibliography

In addition to individual test manuals, the following were used as information sources for this monograph:

Burros, O. Seventh Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1972.

Directory of Developmental Screening Instruments, Minneapolis, Minnesota: Project SEARCH, Minneapolis Public Schools, 1977.

Hoepfner, R., Stern, C. and Nummedal, S.G. (Eds.), CSE-ECRC Preschool/Kindergarten Test Evaluations. Los Angeles: Center for the Study of Evaluation and Early Childhood Research Center, Graduate School of Education, University of California, 1971.

Sattler, J. Assessment of Children's Intelligence. Philadelphia, Pennsylvania: W. B. Saunders Co., 1974.

Tallmadge, Kasten G. The Joint Dissemination Review Panel Ideabook. Washington, D.C., U.S. Dept. of Health, Education and Welfare, October, 1977.

Vane, Julia R. "Problems in and strategies for, evaluating preschool programs." Journal of School Psychology, 1976, Vol. 14, No. 1, pp. 39-46.

FLORIDA EDUCATIONAL
RESEARCH AND DEVELOPMENT COUNCIL, INC.

Officers:

President, Buren L. Dunavant
Treasurer, C. A. Vaughn, Jr.

Board of Directors:

Hugh Flanders, Frank Green, Samuel Hunter, Daniel Hutchison, George A. Marshall, William M. Odom, and Robert Shaffer

Executive Secretary:

Charlie Council, Lee County School Board and Florida Atlantic University Associate Executive Secretary, North Florida Component:
Jeff Pyatte, University of West Florida

Baker County	N. J. Johns
Citrus County	Charles P. Hayes
Collier County	Dr. Gene Barlow
DeSoto County	Jesse A. Blanton
Dixie County	Lloyd Jones
Flagler County	Walter S. Morris, Jr.
Gilchrist County	Leroy Bowdoin
Glades County	Larry Galyon, Supt.
Hamilton County	John Hart
Hernando County	Ruth Marsh
Highlands County	Dr. James Ferrell Smith
Indian River County	Samuel A. Hunter
Lafayette County	Albert J. Winburn, Supt.
Lake County	Dr. C. A. Vaughn, Jr.
Lee County	Hugh C. Flanders
Madison County	Dr. Oscar A. Beck
Manatee County	George A. Marshall
Marion County	Dean Kells
Martin County	Dr. Robert L. Shaffer
Nassau County	Mary Libby
Pasco County	Mary Giella
Polk County	Daniel D. Hutchison
Putnam County	William M. Odom
St. Johns County	Gertrude Taylor
St. Lucie County	James Thomas
Santa Rosa County	R. Frank Green
Sumter County	Wayne Ham
Suwanne County	Everett L. Williamson
Taylor County	Robbins Woodell
Union County	Buren L. Dunavant
Washington County	Kelly V. Brock, Supt.

Associate Membership:

Panhandle Area Educational Cooperative; W. L. Kitching, Dir.